A category-based video-analysis of students' activities in an out-of-school hands-on gene technology lesson

Scharfenberg, Franz-Josef; Bogner, Franz X.; Klautke, Siegfried

A category-based video-analysis of students' activities in an out-of-school hands-on gene technology lesson

Journal:	*International Journal of Science Education*
Manuscript ID:	TSED-2006-0266.R1
Manuscript Type:	Research Paper
Keywords:	biotechnology education
Keywords (user):	video analysis, hands-on experiments, out-of-school learning

Title:

A category-based video-analysis of students' activities in an out-of-school hands-on gene technology lesson

Running head:

Category-based video-analysis of students' lab activities in a gene-technology lab-lesson

Keywords:

Biotechnology education, hands-on experiments, video-analysis, out-of-school learning

Authors:

Franz-Josef Scharfenberg[1], Franz X. Bogner and Siegfried Klautke

Centre of Math & Science Education (Z-MNU),

University of Bayreuth, Institute of Biology Didactics,

Universitätsstr. 30, D-95447 Bayreuth, Germany

[1] To whom correspondence should be sent:

E-Mail: franz-josef.scharfenberg@uni-bayreuth.de

Phone: ++49-921-553640; Fax: ++49-921-552696

TITLE

A category-based video-analysis of students' activities in an out-of-school hands-on gene technology lesson

ABSTRACT

Our research objectives focussed on monitoring (i) students' activities during experimental teaching phases in an out-of-school gene technology lab; (ii) potential relationships with variables such as work group size and cognitive achievement. Altogether, we videotaped 20 work groups of A-level 12th graders (*N* = 67) by continuous recording of their lab-work phases. Subsequent analysis revealed nine categories characterizing the students' most relevant activities. Intra- and inter-observer objectivity as well as reliability scores confirmed the good fit of this categorization. Based on the individual time budgets generated, we extracted four clusters derived from students' prevalent activities. A cross-tabulation of two cluster analysis methods independently used showed a high level of agreement. Clusters were labelled as (i) 'all-rounders' (members of which applied similar portions of time to the main activities), (ii) 'observers' (members' dominating activity focussed on in-group observation of the lab-work), (iii) 'high-experimenters' (members predominantly engaged in specific hands-on activities), and (iv) 'passive students' (members mainly engaged in activities with no experimental relation). Particularly, we found members of clusters 1 and 2 in four-person work groups while members of clusters 3 and 4 were prevalent in three-person groups. During the educational intervention, students of all clusters improved their cognitive achievement on a short-term and a long-term schedule. However, only the 'all-rounders' revealed a high level of persistent (long-term) knowledge with no decrease rate at all. We draw conclusions with respect to work group sizes as well as to organisational aspects of experimental lessons.

Deleted: W

Deleted: -

Deleted: (67

Deleted: [secondary school])

Deleted: M

Deleted: were particularly found

Deleted: C

Deleted: were drawn

INTRODUCTION

Hand-on experiments at school are generally regarded as important for facilitating any learning in science education by providing experiences to students otherwise unavailable in learning science (Hofstein & Lunetta, 2004). On the other hand, experiments in classrooms differ from those in real science. They typically show less variability as well as time and resource limitations (e.g. Füller, 1992). Additionally, in the context of gene technology, legal frameworks may prevent experimentation at school: In Germany, as the law stands, for instance, working with recombinant DNA is usually not allowed in school labs. Universities, museums, science centres, or industrial companies therefore provide dedicated educational laboratories in the field of molecular biology. They typically offer experimental workshops as out-of-school opportunities for hands-on experience (e.g. Maxton-Küchenmeister & Herrmann, 2003). Science education in those labs may also provide authentic hands-on experiences without the typical shortcomings of the classroom (e.g. Markowitz, 2004). In this context, authentic experiments are seen as activities representing 'ordinary day-to-day actions of the community of the practioners' (Hodson, 1998, p. 118). Nevertheless, the gain of such out-of-school experimental lessons is a topic on the current agenda in science education research, although effectiveness of lab work itself frequently has been investigated (e.g. Harlen, 1999; Hofstein & Lunetta, 2004). Additionally, in contrast to physics education where variables such as students' cognitive achievement have been assessed (e.g. Heard, Divall & Johnson, 2000; Semper, 1990) comparable studies of gene technology laboratories do not yet exist.

Deleted: and

Deleted: evince

Deleted: forbidden

Deleted: Euler, 2001

Deleted: observed

Based on this rationale, we offered a daylong teaching unit with experiments conformant with the current syllabus of 12th grade A-level in biology (Bavaria, Germany). Our specific module 'marker genes in bacteria" consisted of a sequence of four experiments:

Deleted: -

(a) transformation of bacteria using a recombinant plasmid coding for the green fluorescent protein (GFP, Tsien, 1998) commonly used as marker protein in molecular biology (e.g. Tromans, 2004);

(b) isolation of the plasmid transformed,

(c) restriction analysis of the plasmid with selected enzymes and

(d) visualisation of students' own results by agarose gel electrophoresis.

All experiments were authentic and followed the criteria of 'authentic inquiry' (Chinn & Malhotra, 2002, p. 118). For instance, students carried out relatively complex controls by testing the survival rates of host bacteria during transformation.

We evaluated our experimental module with regard to cognitive achievement (published elsewhere: Authors, 2006). However, any complete evaluation study also requires 'information and insight about what is really happening when students engage in laboratory activities' (Hofstein & Lunetta, 2004, p. 38). Videotaping is generally seen as an appropriate tool to monitor students' activities during the experimental phases of a teaching unit. (e.g. Seidel, 2005)

Deleted: -

Deleted: , Dalehefte & Meyer

Any lab-work in gene technology education generates new as well as complex situations for the students involved. In a lesson-phase prior to hands-on activities, students have to make predictions about potential experimental results, and after experimentation, they have to match the results to their previously formulated hypotheses. During the experimental phase, they have to read instructions and to operate equipment unknown to them (e.g. variable micropipettes or table centrifuges). Any lack of basic experimental skills might prevent successful lab-work (Bryce & Robertson, 1985) and, in consequence, might decrease any learning outcome. According to authors such as Dunn and Boud (1986), Hodson (1998), and Lunetta (1998), we included a controlled pre-lab exercise in order to assure such basic skills. The pre-lab phase (45 min) consisted of an introduction to the working place by the teacher coupled with students' handling of all relevant equipment. Furthermore, in hands-on phases students generally have to interact and to cooperate within the peer groups accepting specific role requirements within the group work. At least, they have to discuss the procedures to be done and their progress in work. In consequence, experimentation together with others is regarded as a form of cooperative learning (CL) in science teaching (e.g. Tanner, Chatman & Allen, 2003). CL has the potential to provide a framework for promoting scientific process skills (Sherman, 1994). Additionally, better awareness of the objectives to be reached by experimentation may be achieved (Stamovlasis, Dimos & Tsaparlis, 2006). Therefore, according to Johnson, Johnson and Smith (1991) we included some of the important components of CL in the experimental phases:

(a) all resources of a working place had to be shared;

(b) the experimental tasks were generally too difficult to be done individually;

(c) the pre-lab phase gave the group members opportunities to accustom themselves to collaboration, independently of the later tasks ;

(d) we requested participants to change places at the shared working table in order to facilitate learning opportunities to all group members;

(e) we included a variety of pauses in our module, which allowed time for reflection on the processes within the working group.

Any analyses of lab-work phases are challenging due to their complexity (Niedderer, et al., 2002). Any research going beyond this has to apply questionnaires or to observe students directly. Just a few scales have dealt with the context of social interactions, for instance, the Laboratory Interaction Categories (Ogunniyi, 1983) and the Science Laboratory Interaction Categories (Kyle, Penick & Shymansky 1979; Okebukola, 1985). Similarly, a few have dealt with the experimental lesson as a whole, for instance, the Laboratory Analysis Inventory (e.g. Tamir, 1989) originally developed for the analysis of laboratory manuals

(Tamir & Lunetta, 1978) and the Laboratory Program Variables Inventory (Abraham, 1982). Nevertheless, authors have reported three major limitations by using questionnaires (Stigler, Gonzales, Kawanaka, Knoll & Serrano, 1999): (a) different perceptions of complex situations by different persons; (b) low accuracy of questions with regard to given time restrictions; (c) proximity of questionnaires to just a specific set of responses. Direct observations might overcome some of those limitations (e.g. Stigler, et al., 1999). For instance, they might take into account the processes in the learning environment independently of the test persons. Nowadays the state of the art focuses on video surveys including both qualitative and quantitative data collection (Jacobs, et al., 2003). For example, Niedderer et al. (2002) provided a Category-Based Analysis of Videotapes for analysing lab work in physics education. Nevertheless, any categorisation with regard to students' experimentation in the area of gene technology is lacking.

The objectives of our video study were

(a) to categorise students' activities during the experimental phases of a teaching unit;

(b) to explore potential relations of their activity pattern to variables such as prior experiences in experimentation, prior achievement in biology, interest in gene technology, and work group sizes as well as cognitive achievement.

METHODOLOGY

Sample and videotaping

Deleted: -

Altogether, 67 secondary schools students (12th grade, highest stratification level ['Gymnasium']) participated in our video study. All (24 boys, 43 girls, average age 18.0) were enrolled in a regular A-level Biology classroom course ('Leistungskurs'). The sample represents a sub-sample of a main study in our educational laboratory (Scharfenberg, Bogner & Klautke, 2006). All students had already successfully completed a regular half-year A-level

Deleted: Authors

genetics course before participating in our lab-lesson. The videotaped students formed one 2-person, eleven 3-person, and eight 4-person work groups by their own choice. We videotaped all 20 groups in a continuous recording (average duration 51.7 min, *SD* = 9.9) by focal sampling during their experimental phases of the lab-lesson (Martin & Bateson, 1986).

Deleted: -

Deleted: -

Categorisation

We categorized the observed activities of each student during his/her lab-work phases and analysed the individual time budgets (e.g. Figure 1). In principal, our categorisation followed the criteria of content analysis (Bos & Tarnai, 1999), but we partially adopted also relevant results of physics education (e.g. Niedderer, et al. 2002; Seidel, 2005). We employed nine categories for a complete description of the observed activities (Table 1).

[Insert Table 1 here]

We pre-trained two observers in the use of our categorization system, using the software Videograph (Rimmele, 2002). We assessed intra-observer and inter-observer objectivity (Martin & Bateson, 1986) as well as intra-observer and inter-observer reliability (Jacobs, et al., 2003). We randomly selected 3-minutes periods for a second categorization of each videotaped working group. We choose the first period with the lack of the category "activity not visible" for re-categorization by the first observer as well as by the second one. For the objectivity test, we performed a scan sampling every 30 s and compared the chosen categories of the first and the second categorization. As value for intra- and inter-observer objectivity, we computed Cohen's Kappa coefficient (Cohen, 1968; Table 2). According to the criteria of the Video Study within the Third International Mathematics and Science Study 1999 (Jacobs, et al., 2003), we assessed intra- and inter-observer reliability (Table 2).

Deleted: ; Seidel, et al., 2005

Deleted: -

[Insert table 2 here]

[Insert figure 1 here]

Cluster analysis

We used the individual time budgets (e.g. Figure 1) for clustering students based upon similar activity patterns. We extracted a four-cluster solution by an agglomerative hierarchical cluster analysis applying Ward's method (Norusis, 1993). For determining students' cluster membership, we used the K-Means Cluster Analysis procedure (Anderberg, 1973) specifying the cluster number as four. We validated this analysis by a cluster-wise cross-tabulation of the two methods used (Figure 2), revealing a high level of agreement.

[Insert figure 2 here]

According to Bacher (1994), clusters are homogeneous if standard deviations of each variable within each cluster are lower than the corresponding values in the sample as a whole. Almost all our within-cluster values fulfilled this criterion (Table 3).

[Insert table 3 here]

Relations to input and output variables

The procedure of data collection is described elsewhere in detail (Scharfenberg, 2005; Scharfenberg et al., 2006). Prior experiences in experimentation were measured as sum of self-assessment in physics, chemistry, and biology with five categories for specialist assessment due to each science: 0 = *never*, 1 = *seldom*, 2 = *occasional*, 3 = *often*, 4 = *always*; according to Bortz and Döring (1995), this verbal rating has been validated as equidistant. Prior achievement in biology was given as standard of school work in written form. Epistemic interest (Krapp, 2002) in gene technology was rated according to Todt and Götz (2000, nine item scale rated from 0 = *not at all* to 5 = *extraordinary*, Cronbachs Alpha .64). In order to test for students' cognitive achievement at the output side, we applied three different test schedules, a pre-test (T-1) before participation, a post-test (T-2) immediately after the intervention and a retention test (T-3) about six weeks later. We applied nonparametric methods due to existing partially non-normal distribution of variables (Kolmogorov-Smirnov tests (Lilliefors modification): T-1 p = .012; T-2 p = .040; T-3 p = .200, N = 54), and, in

Deleted: Author
Deleted: Authors
Deleted:

consequence, present our results as Boxplots. Changes within all three tests were analyzed with the Friedman test in combination with a pair-wise analysis from T-1 to T-2 and T-3 and from T-2 to T-3 by using Wilcoxon signed-rank test. According to authors such as Bender and Lange (2001), Zöfel (2002), and Diehl and Arbinger (2001), we did not apply possible Bonferroni correction: 'If the global null hypothesis is rejected proceed with 'level α tests for the (...) pair-wise comparison' (Bender & Lange, 2001, p. 1238). Potential differences between the clusters were analysed by the Kruskal-Wallis test.

RESULTS

Cluster analysis

We identified and labelled four clusters with regard to students' prevalent activities: (1) 'all-rounders' (n = 21), (2) 'observers' (n = 17), (3) 'high-experimenters' (n = 15), and (4) 'passive students' (n = 14, Figure 3).

[Insert figure 3 here]

Individual clusters were analysed with regard to single category importance by analysis of variance (ANOVA; Table 4). However, we used the F tests for descriptive purposes only, since the clusters have been chosen to maximize the differences among cases within the four clusters (usually, computed p-values are not corrected for this). Thus, we do not interpret the p-values as tests of hypotheses that the cluster means are equal (Norusis, 1993). For the purpose of a further cluster description, we applied Tuckey tests as post-hoc tests in the cases of significant p-values due to the different sample sizes.

Deleted: n't

[Insert table 4 here]

We labelled cluster-1 as 'all-rounders': whose members distributed their time equally over all relevant activities, 20.2 % for observing purposes, 22.2 % for reading the instruction, 21.2 % for hands-on activities (preparing or reworking as well as experimental steps), and 22.3 % for the different forms of interaction (see Table 4 for details). Especially, their category "activity not visible" dominated all other clusters (Tukey tests: $p \leq .003$ in each case). Cluster-2 was labelled as 'observers', clearly dominating the other clusters by "in-group observing of lab-work" (Tukey tests: $p < .001$ in each case, see Figure 3). Cluster-3 represented the 'high-experimenters', dominating the other clusters their hands-on activities ("preparing or reworking steps" as well as "experimental step", Tukey tests: $p \leq .001$ in each case, see Figure 3). Members of cluster-4 were labelled as 'passive students' characterized by following attributes (see Figure 3): (a) a high proportion of "no experimental relation" (Tukey tests: $p < .001$ in each case); (b) a less one of "reading instruction" and "advising interactions" compared to the 'all-rounders' cluster (Tukey tests: $p = .003$ and $p < .001$); (c) less "out-group interactions" than the 'all-rounders' as well as the 'observers' cluster (Tuckey tests: $p < .001$ in both cases).

Relations to input and output variables

Additionally, we analyzed students' cluster assignment with regard to variables of potential influence on activities during experimental phases: (a) prior experiences with experimentation at school; (b) prior achievement in biology at school; (c) epistemic interest in gene technology, and (d) work group size.

Analysis of variance showed no effect of the first three variables: (a) prior experiences at school, neither with student experiments ($F = 1.107$, $p = .355$, $N = 50$ nor with experiments demonstrated by teachers ($F = 0.921$, $p = .438$, $N = 52$), (b) prior achievement at school ($F = 0.691$, $p=.562$, $N = 52$), and (c) interest in gene technology ($F = 0.709$, $p = .551$, $N = 52$).

Cross-tabulating of group size (d) and cluster assignment revealed a significant relation between these variables (Figure 4). 'All-rounders' and 'observers' dominated the 4-person workgroup while 'high-experimenters' and 'passive students' were prevalent in the 3-person groups. As expected, participants of the only 2-person group were 'high-experimenting students' performing the same hands-on activities as the 3- or 4-person groups as group of two.

[Insert figure 4 here]

Students of all clusters improved their cognitive achievement (T-1 to T-2, Table 5 and Figure 5). In the long-term (T-3), all scores dropped except the 'all-rounders' ones, but never back to the previous levels (T-1). However, we found statistical significance only in the 'high-experimenters' cluster. The lack of significance in the 'observers' and the 'passive students' cluster might originate from the low sample size because the similar change in the sample as a whole was significant (Table 5). Thus, the cluster did not differ in prior knowledge at T-1 (Kruskal-Wallis test χ^2 = 2.737, df = 3, p = .434).

[Insert table 5 here]

[Insert figure 5 here]

DISCUSSION

We focussed in our study on the categorization of students' activities during the experimental phases of a science unit teaching gene technology in a dedicated out-of-school lab. Additionally, we explored students' activity patterns in relation to potential influences such as (at the input side) prior experiences in experimentation, prior achievement at school, epistemic interest in gene technology, and the work group size, or (at the output side) cognitive achievement after the intervention. Our category system extracted specific students' activity patterns in the experimental phases. According to individual students' time budgets four activity types were revealed: 'all-rounders', 'observers', 'high-experimenters', and 'passive students'. 'All-rounders' evenly distributed their time to different types of activities. Their higher value of the category 'activity not visible' is not explainable in detail: Perhaps, 'all-rounder' students were more active compared to the other students thereby obscuring their activities at a higher level. Nevertheless, we do not know what they really did in this part

Deleted: teaching unit

Deleted: than

Deleted: '

of their time (ca. 10 %). It might be possible that someone of this cluster would be assigned to another cluster if his/her activities had been visible. 'Observers' and 'high-experimenters' were clearly characterised by the categories used for labelling. The same is true of the 'passive students' cluster which was primarily labelled as 'passive' because of students' high level of the category 'no experimental relation". However, we see different explanations for this passive behaviour. They might have had a high level of prior experimental experiences, but we did not find any differences between the clusters with regard to prior experiences in experimentation. With regard to the cognitive level, 'passive students' might represent low achievers for whom the content taught was too complex. However, they also might represent high achievers for whom the content taught was previously known and did not challenge them at all. This explanation might fit the critics of CL as failing to benefit low achievers as well high achievers (Slavin, 1984). Nevertheless, we have to take into consideration both that prior achievement at school and specific prior knowledge did not differ in any cluster. Another possible explanation might rise from the social level: 'Passive students' might act just as outsider in their work-groups. However, the cluster did not differ in the level of 'in-group interaction' (Table 4). At least, 'passive students' might have a low level of interest. Although we did not monitor differences in the epistemic interest in gene technology between the four clusters, an object of interest may not only be associated with the content but also with the context and the kinds of activities involved (Hoffmann, 2002). Beside the over-all epistemic part, a lower level of these dimensions may have caused passivity: (i) interest in the specific content of our authentic experiments (as the particular context of the gene technology content generally taught); (ii) interest in the experimentation itself (as the particular activities connected; Gardner, 1985; Hoffmann, 2002). A low level of the categories 'reading instruction' and 'advising interactions' in the 'passive students' cluster (Table 4) may support this hypothesis. Another aspect may rise from the specific effect of the work group size. 'Passive students' dominated the 3-person groups (as well as 'high experimenter' students)

Deleted: these students might generally be regarded as problematic targets of science education and were not found in 4-person groups (Figure 4). Thus, social interactions as well as other cooperative effects within a 4-person group may hinder the 'passive' activity pattern (and reduce the probability of the 'high-experimenter' pattern, too). The specific situation in a gene technology lab may be corroborated by previous and more general results in CL research that 'team formation is most effective when four students work together' (Sherman, 1994, p. 227).

Some researchers have previously assigned roles to students in CL situations. O'Donnell, Dansereau, and Rocklin (1991), for instance, described a 'performer' role (as an active member performing the claimed activity) and a 'listener' role (providing only feedback for his/her partner) within dyads. Horn, Collier, Oxford, Bond, and Dansereau (1998) specifically described a 'learner' and a 'learner facilitator'; the first emerging from a dyad's performer recalls and/or processes information needed for the CL activity; the latter serves as explainer as well as provider of supportive material. Additionally, Wenzel (2000) described the 'leadership' role. Such a student takes initiatives for answering and/or explanations thereby leading interactions (Stamovlasis, Dimos & Tsarpalis, 2006), but he/she may occasionally monopolise work and lead to dysfunctional results of CL (Wenzel, 2000). Nevertheless, only the latter role of leadership might fit to one of the four clusters given by our experimental situation. The 'high-experimenter' students clearly dominated their work group's hands-on activities thus resembling a leading function, and maybe monopolising hands-on activities, too. Stamovlasis et al. (2006, p. 562) assigned three roles, an 'active member', a 'very active member', and a 'spectator" based on utterance analysis as measure of individual student's involvement in CL interactions. The 'active member' may comply the 'all-rounder' found in our study. The 'very active member' is characterised by a higher contribution to the 'total number of utterances' than the average and the 'spectator' by a lower contribution to the 'total number of utterances'. The 'very active member' may comply with our 'high-experimenter' while the 'spectator' may comply with our 'passive student' in our hands-on situation. In contrast to the both active members, the 'spectators' are 'did not profit from CL (…) as a result of non-participation in the process of CL' (p. 567). A match with our active 'observer' failed to emerge; perhaps such behaviour may just emerge only in real hands-on settings. All three roles showed a decrease in the long term knowledge (Stamovlasis et al. 2006) while our 'all-rounders' did not loose their acquired knowledge again, maybe caused by the more active situation in our hands-on setting. Thus, an 'all-rounder' contributing his/her time more or less equally to the different activities during experimentation may learn best in a hands-on group work.

A limitation of our video study is the lack of a transcription analysis with regard to students' verbalising. Although such an analysis might be ideal as an additional step in science education research, our aim only was the analysis on the level of activities. Further research may connect this level with the level of the specific talks during experimentation.

CONCLUSIONS

With regard to experimentation embedded in a complex content, our video study may indicate advantages for teaching: (i) Portioning the time available in similar fashion to hands-on activities might benefit students at the cognitive level. (ii) Assigning students to 4-person work groups might be useful in order to prevent a potential 'passive' behaviour.

Deleted: prevent to favour

Teacher might guide their students towards 'all-rounder' behaviour. A consistent movement at the working place rather than sitting as in the classroom may provide a supportive environment. A single request for that at the beginning of a module (as we did) seems to be insufficient. Furthermore, the teacher might especially address the desired 'all-rounder' behaviour type to the students. Prior to the experimentation phase, he/she might request an equal distribution of the given activities to all students of the work group. A more convincing method for the students might be the use of cooperation scripts. They clearly describe what has to be done coupled with personal addressing of the working steps. The purpose of such socio-cognitive scripts has already been shown in e-learning settings (e.g.

Ertl, Fischer & Mandl, 2006; Reynolds, Patterson, Skaggs & Danserau, 1991). These scripts resemble to a specific role assignment prior to the activities. However, Chang and Lederman (1994, p.169) tested the effect of a preliminary assignment of more global roles (in this case, 'manager, investigator, and recorder') to the students in laboratory classes and found no significant effect on their achievement. Assigning students to 4-person groups might be an uncomplicated solution. Especially in complex hands-on situation in an out-of-school learning environment, one might convince class teachers to employ this work group size. Further research will possibly reveal potential advantages of using the described cooperative scripts in hands-on science education. Another approach we follow is the request of more content specific interactions within the work group. We intend to introduce a group discussion phase prior to the hands-on phase coupled with writing down students' content specific ideas. We expect thus to reduce the amount of 'passive' behaviour.

Deleted: to

Deleted:

ACKNOWLEDGEMENTS

We welcome the cooperation of teachers and students involved in this study as well as we appreciate the helpful and valuable discussion of the manuscript with M. Wiseman. The study was funded by Bavarian State Ministry of the Environment, Public Health and Consumer Protection, Bavarian State Ministry of Education, and German Science Foundation (DFG KL 664/5-1 and BO 944/4-2).

REFERENCES

Abraham, M. R. (1982). A descriptive instrument for use in investigating science laboratories. *Journal of Research in Science Teaching*, 19 (2), 155-165.

Anderberg, M. R. (1973). *Cluster Analysis for Applications*. New York: Academic Press.

Bacher, J. (1994). *Clusteranalyse. Anwendungsorientierte Einführung* [Cluster analysis. Application-oriented introduction]. München, Germany: R. Oldenbourg.

Bender, R., & Lange, S. (2001). Adjusting for multiple testing - when and how? *Journal of Clinical Epidemiology*, 54, 343-349.

Bortz, J., & Döring, N. (1995). *Forschungsmethoden und Evaluation* (2nd ed.) [Methods of research and evaluation]. Berlin, Germany: Springer.

Bos, W., & Tarnai, C. (1999). Content analysis in empirical research. *International Journal of Educational Research*, 31 (6), 659-671.

Bryce, T., & Robertson, I. (1985). What can they do? A review of practical assessment in Science. *Studies in Science Education*, 12, 1-24.

Chang, H.-P., & Lederman, N. (1994). The effect of levels of cooperation within physical science laboratory groups on physical science achievement. *Journal of Research in Science Teaching*, 31 (2), 167-181.

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, 86 (2), 175-218.

Cohen, J. (1968). Weighted Kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70, 213-220.

Diehl, J. M., & Arbinger, R. (2001). *Einführung in die Interferenzstatistik* (3rd ed.) [Introduction in interference statistics]. Eschborn, Germany: Dietmar Klotz.

Dunn, J., & Boud, D. (1986). Sequencing and organization. In Boud, D., Dunn, J., & Hegarty-Hazel, E. (Eds.) *Teaching in Laboratories* (pp. 57-78). Exeter, UK: NFER-Nelson

Ertl, B., Fischer, F., & Mandl, H. (2006). Conceptual and socio-cognitive support for collaborative learning in videoconferencing environments. *Computers & Education*, 47 (3), 298-315.

Füller, F. (1992). *Biologische Unterrichtsexperimente: Bedeutung und Effektivität* [Biological experiments at school: Importance and efficacy]. München, Germany: Münchner Schriften zur Didaktik, Vol. 8.

Deleted: Euler, M. (2001). Lernen durch Experimentieren [Learning by experimenting]. In Ringelband, U., Prenzel, M., & Euler, M. (Eds.) *Lernort Labor - Initiativen zur naturwissenschaftlichen Bildung zwischen Schule, Forschung u. Wirtschaft* [Out-of-school labs – initiatives for science education by schools, science, and economy]. Kiel, Germany: IPN.¶

Harlen, W. ,(1999). *Effective teaching of science. A review of research*. Edinburgh, UK: SRCE Publication.

Heard, P. F., Divall, A., & Johnson, S. D. (2000). Can 'ears-on' help hands-on science learning - for girls and boys? *International Journal of Science Education*, 18 (11), 1133-1146.

Hill, G. (1982). Group versus individual performance: Are N + 1 heads better than one. *Psychological Bulletin*, 19 (3), 517-539.

Hodson, D. (1998). *Teaching and Learning Science. Towards a personalized approach*. Buckingham, UK: Open University Press.

Hoffmann, L. (2002). Promoting girls' interest and achievment in physics classes for beginners. *Learning and Instruction*, 12 (4), 447-465.

Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education*, 88 (1), 28-54.

Horn, E., Collier, W., Oxford, J., Bond, C., & Dansereau, D. (1998). Individual differences in dyadic cooperative learning. *Journal of Educational Psychology*, 90 (1), 153-161.

Jacobs, J., Garnier, H., Gallimore, R., Hollingsworth, H., Bogard Givvin, K., Rust, K. et al. (2003). *Third International Mathematics and Science Study 1999 Video Study Technical Report Volume 1: Mathematics Technical Report*. Washington DC: National Center for Education Statistics, U.S. Department of Education.

Johnson, D., Johnson, R., & Smith, K. (1991). *Active learning: Cooperation in the college classroom*. Edina, MN: Interaction Book Company.

Krapp, A. (2002). Structural and dynamic aspects of interest development: theoretical considerations from an ontogenetic perspective. *Learning and Instruction*, 12 (4), 383-409.

Kyle, W. C. J., Penick, J., & Shymansky, J. (1979). Assessing and analysing the performance of students in college science laboratories. *Journal of Research in Science Teaching*, 16 (6), 545-551.

Lunetta, V. N. (1998). The school science laboratory: Historical perspectives and contexts for contempory teaching. In Fraser, B. J., & Tobin, K. J. (Eds.), *International Handbool of Science Education; Part One* (pp. 249-262). Dordrecht, Netherlands: Kluwer Academic.

Markowitz, D. G. (2004). Evaluation of the long-term impact of a university high school summer science program on students' interest and perceived abilities in science. *Journal of Science Education and Technology*, 13 (3), 395-407.

Martin, P., & Bateson, P. (1988). *Measuring behaviour. An introductory guide* (2^{nd} ed.). Cambrigde, UK: Cambrigde University Press.

Maxton-Küchenmeister, J., & Herrmann, R. (2003). Genlabor & Schule - eine Übersicht über Experimentierangebote zur Vermittlung von Gen- und Biotechnologie an Schulen [Out-of-school labs and school – an overview about offers with regard to gene and biotechnology]. *BIOspektrum*, 9 (4), 382-385.

Niedderer, H., Aufschaiter, S. v., Tiberghien, A., Haller, K., Hucke, L., Sander, F., et al. (2002): Talking physics in labwork contexts - A category-based analysis of videotapes. In Psillos, D., & Niedderer, H. (Eds.) *Teaching and learning in the science laboratory* (pp. 31-40). Dordrecht, Netherlands: Kluwer Academic.

Deleted: i

Norusis, M. J. (1993). *SPSS for Windows Professional Statistics Release 6.0*. Chicago: SPSS Inc.

O'Donnell, A., Danserau, D., & Rocklin, T. (1991). Individual differences in the cooperative learning of concrete procedures. *Learning and Individual Differences*, 3 (2), 149-162.

Ogunniyi, M. (1983). An analysis of laboratory activities in selected Nigerian secondary schools. *European Journal of Science Education*, 5 (2), 195-201.

Okebukola, P. (1985). Science laboratory behaviour strategies of students relative to performance in and attitude to laboratory work. *Journal of Research in Science Teaching*, 22 (3), 221-232.

Reynolds, S., Patterson, M., Skaggs, L., & Dansereau, D. (1991). Knowledge hypermaps and cooperative learning. *Computers Education*, 16 (2), 167-173.

Rimmele, R. (2002). *Das Programm Videograph* [The software Videograph]. Kiel, Germany: IPN.

Scharfenberg, F.-J. (2005). *Experimenteller Biologieunterricht zu Aspekten der Gentechnik im Lernort Labor: empirische Untersuchung zu Akzeptanz, Wissenserwerb und Interesse* [Experimental gene technology teaching: empirical analysis with regard to acceptance, knowledge achievement and interest]. PhD-Thesis, University of Bayreuth, Germany.

Scharfenberg, F.-J., Bogner, F., & Klautke, S. (2006). Learning in a gene technology lab with educational focus: Results of a teaching unit with authentic experiments. *Biochemistry and Molecular Biology Education*, in press.

Seidel, T. (2005). Video analysis strategies of the IPN Video Study - a methodological overview. In Seidel, T., Prenzel, M., & Kobarg, M. (Eds.) *How to run a video study*. New York: Waxmann.

Semper, R. J. (1990). Science museums as environments for learning. *Physics Today*, 43 (November), 50-56.

Deleted: Seidel, T., Dalehefte, I. M., & Meyer, L. (2005). Standardized guidelines - How to collect videotapes. In Seidel, T., Prenzel, M., & Kobarg, M. (Eds.) *How to run a video study*. New York: Waxmann.¶

Sherman, S. (1994). Cooperative learning and science. In Sharan, S. (Ed.) Handbook *of cooperative learning methods*. Westport, CT:Greenwood Press.

Slavin, R. (1984). Meta-analysis in education: How has it been used. *Educational Researcher*, 13, 6-15.

Stamovlasis, D., Dimos, A., & Tsaparlis, G. (2006). A study of group interaction processes in learning lower secondary physics. *Journal of Research in Science Teaching*, 43 (6), 556-576.

Stigler, J., Gonzales, P., Kawanaka, T., Knoll, S., & Serrano, A. (1999). *The TIMSS videotape classroom study: Methods and findings from an exploratory research project on eight-grade mathematics instruction in Germany, Japan,and the United States.* Washington, DC: U.S. Government Printing Office.

Tamir, P. (1989). Training teachers to teach effectively in the laboratory. *Science Education*, 73 (1), 59-69.

Tamir, P., & Lunetta, V. (1978). An analysis of laboratory inquiries in the BSCS yellow version. *The American Teacher*, 40 (September), 353-357.

Tanner, K., Chatman, L., & Allen, D. (2003). Approaches to cell biology teaching: Cooperative learning in the science classroom - beyond students working in groups. *Cell Biology Education*, 2 (spring), 1-5.

Todt, E., & Götz, C. (2000). Interests and attitudes of adolescents regarding genetic engineering. In Bayrhuber, H., Garvin, W., & Grainger, J. (Eds.) *Teaching biotechnology at school: A European perspective*. Kiel, Germany: IPN.

Tromans, A. (2004). Bright future for GFP. *Nature Reviews Molecular Cell Biology*, 5 (November), 865.

Tsien, R. Y. (1998). The green fluorescent protein. *Annual Review of Biochemistry*, 67, 509-544.

Wenzel, T. (2000). Cooperative students' activities as learning devices. *Analytical Chemistry A*, 72 (April 1), 293A-296A.

Zöfel, P. (2002): *Statistik verstehen. Ein Begleitbuch zur computergestützten Anwendung* [Understanding statistics. A book for computer-aided applications]. München, Germany: Addison-Wesley.

TABLE 1: Categories for the description of students' observed activities during their lab-work phases

Category	*Description*	*Example*
Activity not visible	S's activity cannot be seen because he/she is hidden behind another person or by equipment or he/she has left the working place.	The teacher stands in front of S.
No experimental relation[a]	S shows an activity not related to the experimental phase.	S phones with his/her mobile.
Out-group interaction[b]	S contacts a S of another workgroup.	S turns to S of the workgroup behind.
Advising interaction[b]	S contacts an adviser either the lab teacher or his/her present own teacher or the present assistant.	S asks the assistant.
In-group interaction[b]	S contacts S of his/her own workgroup.	S talks to S beside.
In-group observing of lab-work[c]	S looks at experimental steps as well as preparing or reworking steps done by S of his/her own workgroup by visible turning towards and without doing anything else.	S looks at S beside doing a lab-work step.
Reading instruction[d]	S reads the written instruction.	S looks at his instruction
Preparing or reworking steps[e]	S prepares an experimental step or finishes it by a reworking step.	S adjusts a graduated pipette.
Experimental step[f]	S performs an experimental step written in the instruction.	S pipettes a given volume.

Note: Abbreviations used in following tables and figures: [a]No exper. relation; [b]interact.; [c]In-group observ.; [d]Read. instruction; [e]Prepar. or rework.; [f]Exper. steps.

TABLE 2: Intra- and inter-observer objectivity and reliability of categorization

Observer	Objectivity (Cohens Kappa[a])		Reliability (Percentage of concordant coded phases[b])	
	Intra-observer	Inter-observer	Intra-observer	Inter-observer
1	.82	.71	.88	.83
2	.80	.69	.88	.75

Notes: [a] Wolf (1997, p. 964) assesses kappa values between .41 and .60 as ‘moderate’, between .61 and .80 as ‘substantial’ and > .80 as ‘almost perfect’.

[b] We rated every matching in-point and out-point of a coding interval within a phase of maximally ten seconds as well as every concordantly coded phase between these points as consistently coded time phase (Jacobs, et al., 2003).

TABLE 3: Analysis of cluster homogeneity by comparison of the standard deviations of each variable within the clusters with the values in the sample as a whole[a]

Category	Cluster (N)				All (N)
	1	2	3	4	
	(21)	(17)	(15)	(14)	(67)
Activity not visible	4,61[a]	2,59	3,59	3,80	4,44
No experimental relation	2,96	6,32	4,25	5,48	9,61
Out-group interaction	3,25	4,61	2,57	1,93	4,78
Advising interaction	1,92[a]	1,30	1,39	1,58	1,80
In-group interaction	2,45	3,92	5,67[a]	3,73	4,09
In-group observing of lab-work	3,53	3,85	4,06	3,66	8,10
Reading instruction	5,68	5,74	5,12	5,50	5,96
Preparing or reworking steps	3,50	2,79	6,75	4,12	6,94
Experimental steps	2,95	3,22	3,17	3,09	3,85

Note: [a] Only three (of 36, grey background) within-cluster values showed a higher level as the corresponding values in the sample as a whole.

TABLE 4: Cluster specific ANOVA of the observed activities

Category	Percentage	of time	budget	(*M*)	ANOVA	
	Cluster	(*N*)			*F*	*p*
	All-rounders (21)	Observers (17)	High-experimenters (15)	Passive students (14)		
Activity not visible	10.2	5.7	5.0	4.3	9.38	<.001
No experimental relation	3.8	6.2	10.3	26.1	66.71	<.001
Out-group interaction	7.8	10.1	2.4	1.7	24.73	<001
Advising interaction	4.9	4.3	3.8	2.5	6.88	<001
In-group interaction	9.6	7.3	10.5	7.9	2.31	.085
In-group observing of lab-work	20.2	31.0	11.0	16.4	81.02	<.001
Reading instruction	22.2	18.5	19.5	15.2	4.51	.006
Preparing or reworking steps	13.2	11.4	25.5	18.7	33.34	<.001
Experimental steps	8.0	5.4	12.1	7.3	12.95	<.001

Note: ANOVA was performed based on following prerequisites: Variables were normally distributed (Kolmogorov-Smirnov tests [Lilliefors modification]: $p > .051$ in each case, $N = 67$). Homogeneity of variance was given for six variables (Levene tests: $p > .073$ in each case). Nevertheless, we accepted the three significant values (Levene tests: $p < .029$ in each case) by adjusting the *p*-value for significance for the analysis of variance to .01 (Zöfel 2002).

TABLE 5: Changes in scoring of knowledge in the four clusters as well as in the sample as a whole

Test dates	Changes of medians (grouped)				
	All-rounders (N=17)	Observers (N=16)	High-experimenters (N=10)	Passive students (N=11)	All (N=54)
T-1 / T-2	6.0 to 9.8 ***	4.4 to 9.3 **	4.5 to 10.8 **	5.5 to 10.0 **	5.0 to 10.0 ***
T-2 / T-3[a]	9.8 to 9.7	9.3 to 6.8	10.8 to 8.0 *	10.0 to 8.2	10.0 to 8.2 ***

Notes: significant differences * $p < .05$, ** $p < .01$; *** $p < .001$ (for statistical tests see Appendix).

[a] All changes from T-1 to T-3 were significant, too (see Appendix).

FIGURE CAPTIONS

FIGURE 1: Example of a student's individual time budget showing his/her categorised activities during the experimental phases.

FIGURE 2: Validation of the cluster analysis by cluster-wise cross-tabulation of both methods used: Ward's method and K-Means procedure (coefficient of contingency C = .83 with C_{max} = .87, $N = 67$, $p < .001$).

FIGURE 3: Characterization of the four clusters identified with regard to students' activities during the experimental phases (see text for details).

FIGURE 4: Significant relation between work group size and cluster assignment (coefficient of contingency $C = .64$ with estimated $C_{max} = .85$, $N = 67$ and $p < .001$).

FIGURE 5: Changes in knowledge scores in the four clusters over the three test schedules.

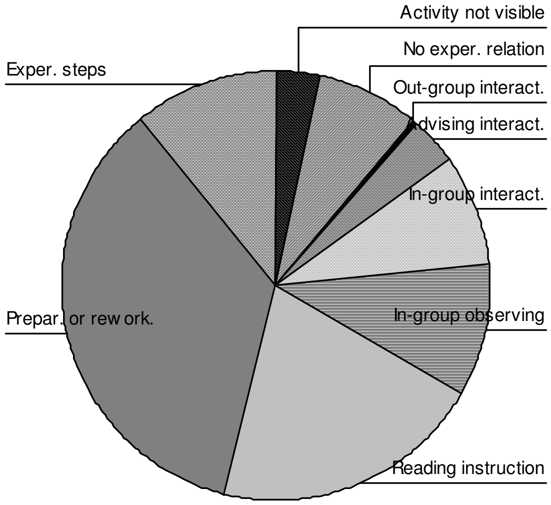

Activity not visible
No exper. relation
Out-group interact.
Exper. steps
dvising interact.
In-group interact.
In-group observing
Prepar. or rew ork.
Reading instruction

FIGURE 1

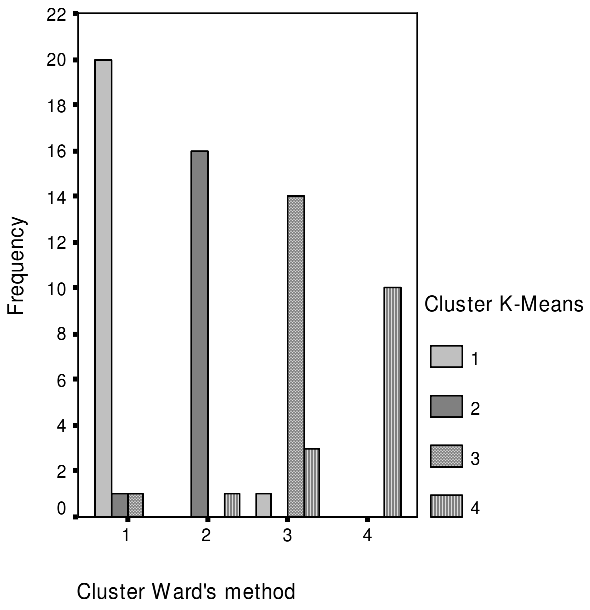

22
20
18
16
14
12
10
8
6
4
2
0
Frequency
1
2
3
4
Cluster Ward's method
Cluster K-Means
1
2
3
4

FIGURE 2

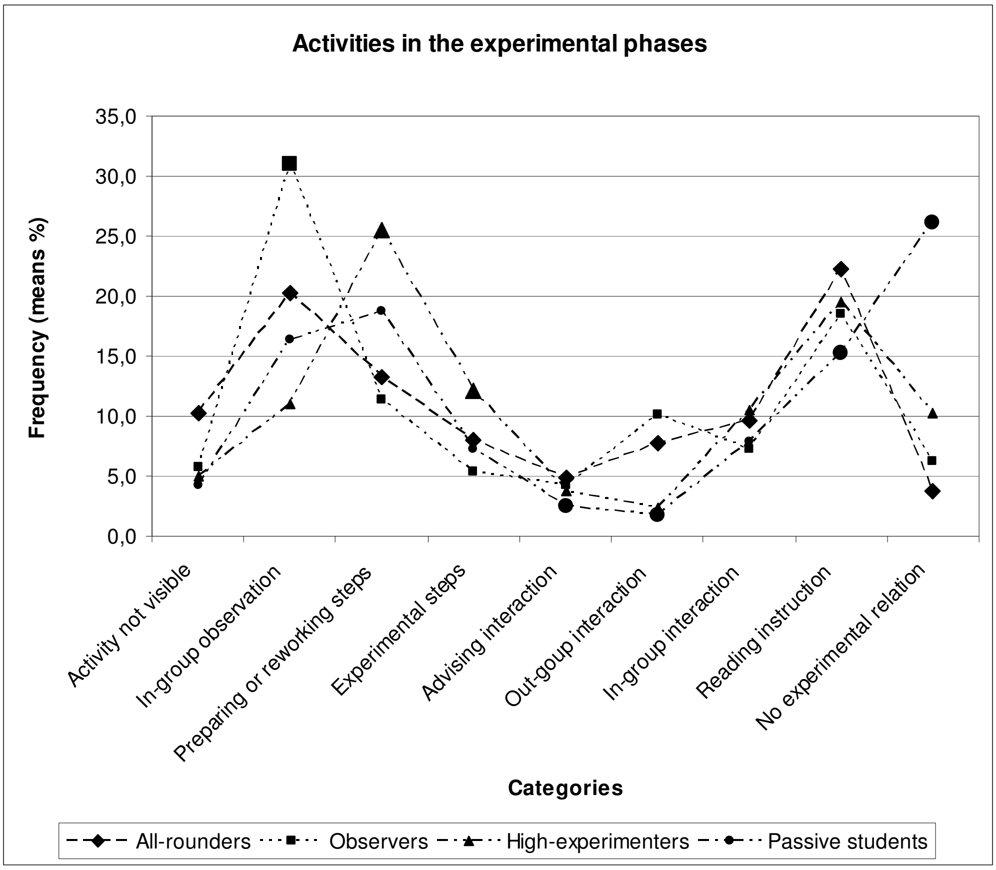

Activities in the experimental phases
35,0
30,0
25,0
20,0
15,0
10,0
5,0
0,0
Frequency (means %)
Activity not visible
In-group observation
Preparing or reworking steps
Experimental steps
Advising interaction
Out-goup interaction
In-group interaction
Reading instruction
No experimental relation
Categories
All-rounders
Observers
High-experimenters
Passive students

FIGURE 3

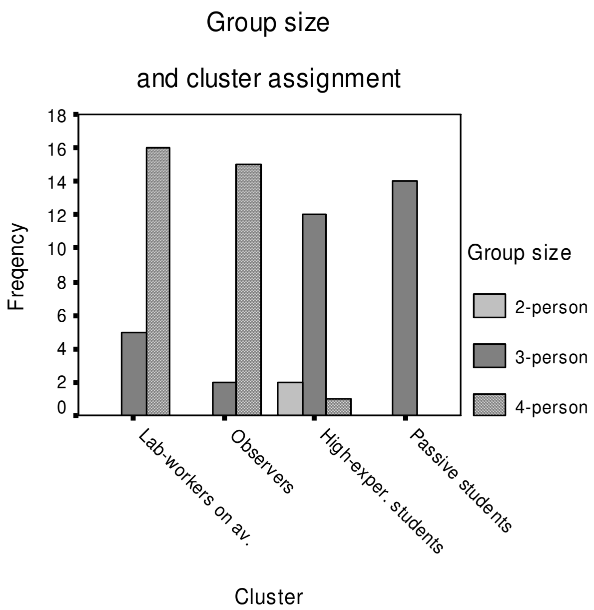
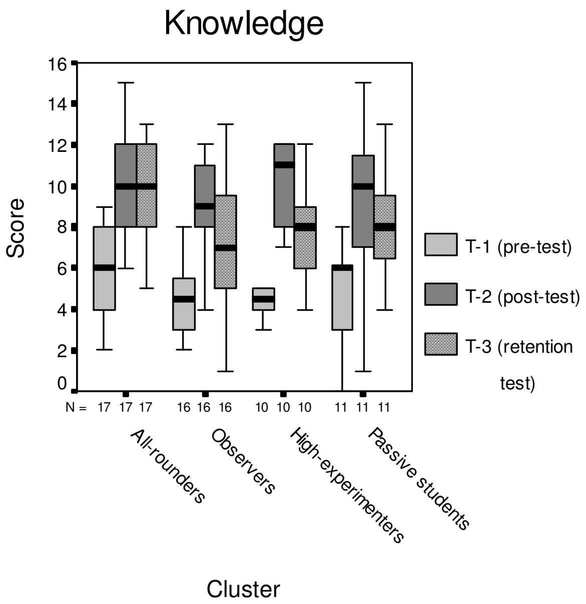

Group size
and cluster assignment
Freqency
18
16
14
12
10
8
6
4
2
0
Group size
2-person
3-person
4-person
Lab-workers on av.
Observers
High-exper. students
Passive students
Cluster

FIGURE 4



Knowledge
Score
0
2
4
6
8
10
12
14
16
N = 17 17 17 16 16 16 10 10 10 11 11 11
All-rounders
Observers
High-experimenters
Passive students
Cluster
T-1 (pre-test)
T-2 (post-test)
T-3 (retention test)

FIGURE 5

APPENDIX: Friedman tests (in any case df=2) and subsequent pair-wise Wilcoxon tests demonstrate cognitive achievement in the students' sample as a whole (N=54) as well as specific results of the four clusters

Group		Friedman test		Wilcoxon signed-rank test					
				T-1 / T-2		T-2 / T-3		T-1 / T-3	
	N	Chi-Square	*p*	Z	*p*	Z	*p*	Z	*p*
Whole sample	54	72.07	<.001	-6.27	<.001	-3.59	<.001	-5.91	<.001
Cluster-1[a]	17	24.29	<.001	-3.63	<.001	-1.42	.153	-3.20	.001
Cluster-2[b]	16	16.26	<.001	-3.32	.001	-1.82	.068	-3.08	.002
Cluster-3[c]	10	14.7	.001	-2,81	.005	-2.31	.020	-2.56	.01
Cluster-4[d]	11	18.42	<.001	-2.82	.005	-1.83	.067	-2.95	.003

Note: [a]'All-rounders'; [b]'observers'; [c]'high-experimenters'; [d]'passive students'.